| | Application No. | | Applicant(s) | |
|---|---|---|---|---|
| **Notice of Allowability** | 10/603,035 | | CHAUDHURI ET AL. | |
| | Examiner | | Art Unit | |
| | DIANE D. MIZRAHI | | 2165 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--*

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to *8-30-06*.

2. ☒ The allowed claim(s) is/are *1-46*.

3. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☐ All    b) ☐ Some*    c) ☐ None    of the:

        1. ☐ Certified copies of the priority documents have been received.

        2. ☐ Certified copies of the priority documents have been received in Application No. _____ .

        3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the

            International Bureau (PCT Rule 17.2(a)).

    * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application. **THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.**

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.

5. ☐ CORRECTED DRAWINGS ( as "replacement sheets") must be submitted.

    (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review ( PTO-948) attached

        1) ☐ hereto or 2) ☐ to Paper No./Mail Date _____ .

    (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of

        Paper No./Mail Date _____ .

    Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

**Attachment(s)**

1. ☐ Notice of References Cited (PTO-892)

2. ☐ Notice of Draftperson's Patent Drawing Review (PTO-948)

3. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____

4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material

5. ☐ Notice of Informal Patent Application

6. ☒ Interview Summary (PTO-413), Paper No./Mail Date *8-30-06* .

7. ☐ Examiner's Amendment/Comment

8. ☐ Examiner's Statement of Reasons for Allowance

9. ☐ Other _____ .

DIANE D. MIZRAHI
PRIMARY EXAMINER

## EXAMINER'S AMENDMENT

An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Attorney Stephen Siu August 30, 2006.

### *The application has been amended as follows:*

1. (Currently amended) A method of estimating selectivity of a given string predicate of length n in a database query, comprising:

a) estimating selectivities of a plurality of string predicate substrings, the plurality of string predicate substrings including substrings of the given string predicate and having each substring length between q to n, where q < n;

b) categorizing each of the string predicate substrings based on length;

c) selecting one candidate substring for each category of substring length based on estimated selectivities of the substrings to obtain a plurality of candidate identifying substrings, each candidate identifying substring in the plurality of identifying substrings having a different length between q and n;

d) combining the estimated selectivities of each of the candidate substrings in the plurality of identifying substrings; and

e) ~~returning~~ generating the combined estimated
selectivities of the candidate substrings as the estimated
selectivity of the given string predicate.

2. (Original) The method of claim 1 further comprising storing
selectivity information for the database and using stored
selectivity information to estimate the selectivities of the
substrings of various lengths.

3. (Original) The method of claim 1 wherein a substring with a
lowest estimated selectivity is selected as the candidate
substring at each length.

4. (Original) The method of claim 1 further comprising
calculating exact selectivities of substrings of a given maximum
length and using the exact selectivities to estimate the
selectivities of the substrings of various substring lengths.

5. (Original) The method of claim 4 wherein a range of the
various substring lengths whose selectivities are estimated is
between the given maximum length of the substrings whose
selectivities are calculated exactly and the length of the given
string predicate.

6. (Original) The method of claim 4 wherein the candidate
substring for the length equal to the given maximum length of
the substrings whose selectivities are calculated exactly is
selected based on the exact selectivity of the substring.

7. (Original) The method of claim 1 wherein a q-gram table is constructed for substrings of a given maximum length and is accessed to estimate selectivities of substrings of various substring lengths.

8. (Original) The method of claim 4 wherein a markov estimator uses the exact selectivities to estimate the selectivities of the substrings of various substring lengths.

9. (Original) The method of claim 1 wherein characteristics of string values in a relation of the database are used to combine the estimated selectivities of the candidate substrings.

10. (Original) The method of claim 1 wherein characteristics of a workload of queries are used to combine the estimated selectivities of the candidate substrings.

11. (Original) The method of claim 1 wherein a model for combining the estimated selectivities of candidate substrings is learned from query workloads.

12. (Original) The method of claim 1 wherein said model is applied to the candidate substrings at run time to estimate the string predicate selectivity.

13. (Original) The method of claim 1 wherein the given string predicate is a unit predicate.

14. (Original) The method of claim 1 wherein the given string predicate includes a wildcard character.

15. (Original) The method of claim 1 wherein the given string predicate is a range predicates.

16. (Original) The method of claim 1 wherein weights are assigned to each length of candidate substring to combine the selectivities of the candidate substrings.

17. (Original) The method of claim 16 wherein a function for assigning said weights is learned from data sets of the database.

18. (Original) The method of claim 16 wherein a function for assigning said weights is learned from an expected query workload.

19. (Original) The method of claim 16 further comprising calculating actual selectivities of substrings of queries from an expected workload and determining estimated selectivities of the substrings of a queries from the expected workload to learn a function for assigning said weights.

20. (Original) The method of claim 16 further comprising calculating for a string predicate of a query from an expected workload an actual selectivity of a candidate substring having the given length, determining for the string predicate of the query from the expected workload an estimated selectivity of the candidate substring having the given length, and assigning a weight to candidate substrings of a given length by based on a

relationship between the calculated actual selectivity and the determined estimated selectivity.

21. (Original) The method of claim 1 wherein selectivities of the candidate substrings are combined using regression trees.

22. (Original) The method of claim 20 wherein said regression trees are learned from data sets of the database.

23. (Original) The method of claim 20 wherein said regression trees are learned from an expected query workload.

24. (Currently amended) A computer readable storage medium having computer executable instructions stored thereon for performing a method of estimating selectivity of a given string predicate of length n in a database query, the method comprising:

   a) estimating selectivities of a plurality of substrings, the plurality of string predicate substrings including substrings of the given string predicate and having each substring length between q to n, where q < n;

   b) categorizing each of the string predicate substrings based on length;

   c) selecting one candidate substring for each category of substring length based on estimated selectivities of the substrings to obtain a plurality of candidate identifying substrings, each candidate identifying substring in the plurality of identifying substrings having a different length between q and n;

d) combining the estimated selectivities of each of the candidate substrings in the plurality of identifying substrings; and

e) ~~returning~~ generating the combined estimated selectivities of the candidate substrings as the estimated selectivity of the given string predicate.

25. (Currently amended) The computer readable storage medium of claim 24 wherein the method further comprises storing selectivity information for the database and using stored selectivity information to estimate the selectivities of the substrings of various lengths.

26. (Currently amended) The computer readable storage medium of claim 24 wherein a substring with a lowest estimated selectivity is selected as the candidate substring at each length.

27. (Currently amended) The computer readable storage medium of claim 24 wherein the method further comprises calculating exact selectivities of substrings of a given maximum length and using the exact selectivities to estimate the selectivities of the substrings of various substring lengths.

28. (Currently amended) The computer readable storage medium of claim 27 wherein a range of the various substring lengths whose selectivities are estimated is between the given maximum length of the substrings whose selectivites are calculated exactly and the length of the given string predicate.

29. (Currently amended) The computer readable storage medium of
claim 27 wherein the candidate substring for the length equal to
the given maximum length of the substrings whose selectivities
are calculated exactly is selected based on the exact
selectivity of the substring.

30. (Currently amended) The computer readable storage medium of
claim 24 wherein a q-gram table is constructed for substrings of
a given maximum length and is accessed to estimate selectivities
of substrings of various substrings lengths.

31. (Currently amended) The computer readable storage medium of
claim 28 wherein a markov estimator uses the exact selectivities
to estimate the selectivities of the substrings of various
substring lengths.

32. (Currently amended) The computer readable storage medium of
claim 24 wherein characteristics of string values in a relation
of the database are used to combine the estimated selectivities
of the candidate substrings.

33. (Currently amended) The computer readable storage medium of
claim 24 wherein characteristics of a workload of queries are
used to combine the estimated selectivities of the candidate
substrings.

34. (Currently amended) The computer readable storage medium of
claim 24 wherein a model for combining the estimated
selectivities of candidate substrings is learned from query
workloads.

35. (Currently amended) The computer readable storage medium of claim 24 wherein said model is applied to the candidate substrings at run time to estimate the string predicate selectivity.

36. (Currently amended) The computer readable storage medium of claim 24 wherein the given string predicate is a unit predicate.

37. (Currently amended) The computer readable storage medium of claim 24 wherein the given string predicate includes a wildcard character.

38. (Currently amended) The computer readable storage medium of claim 24 wherein the given string predicate is a range predicates.

39. (Currently amended) The computer readable storage medium of claim 24 wherein weights are assigned to each length of candidate substring to combine the selectivities of the candidate substrings.

40. (Currently amended) The computer readable storage medium of claim 39 wherein a function for assigning said weights is learned from data sets of the database.

41. (Currently amended) The computer readable storage medium of claim 39 wherein a function for assigning said weights is learned from an expected query workload.

42. (Currently amended) The computer readable storage medium of claim 39 wherein the method further comprises calculating actual selectivities of substrings of queries from an expected workload and determining estimated selectivities of the substrings of a queries from the expected workload to learn a function for assigning said weights.

43. (Currently amended) The computer readable storage medium of claim 39 wherein the method further comprises calculating for a string predicate of a query from an expected workload an actual selectivity of a candidate substring having the given length, determining for the string predicate of the query from the expected workload an estimated selectivity of the candidate substring having the given length, and assigning a weight to candidate substrings of a given length by based on a relationship between the calculated actual selectivity and the determined estimated selectivity.

44. (Currently amended) The computer readable storage medium of claim 24 wherein selectivities of the candidate substrings are combined using regression trees.

45. (Currently amended) The computer readable storage medium of claim 44 wherein said regression trees are learned from data sets of the database.

46. (Currently amended) The computer readable storage medium of claim 44 wherein said regression trees are learned from an expected query workload.

### *Allowable Subject Matter*

Claims 1-46 are allowed over the prior art made of record.

### *Comments*

The date on which the petition under 37 CFR 1.136(a) and the appropriate extension fee have been filed is the date for purposes of determining the period of extension and the corresponding amount of the fee. In no case may an applicant reply outside the SIX (6) MONTH statutory period or obtain an extension for more than FIVE (5) MONTHS beyond the date for reply set forth in an Office action. A fully responsive reply must be timely filed to avoid abandonment of this application.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

As allowable subject matter has been indicated, Applicant's response must either comply with all formal requirements or specifically traverse each requirement not complied with. See 37 CRF 1.111(b) and MPEP section 707.07(a).

### *Other Prior Art Made of Record*

The prior art made of record and not relied upon is considered pertinent to Applicant's disclosure.  U.S. patents and U.S. patent application publications will not be supplied with Office actions.   Examiners advises the Applicant that the cited U.S. patents and patent application publications are available for download via the Office's PAIR.  As an alternate source, all U.S. patents and patent application publications are available on the USPTO web site (www.uspto.gov), from the Office of Public Records and from commercial sources.  For the use of the Office's PAIR system, Applicants may refer to the Electronic Business Center (EBC) at http://www.uspto.gov/ebc/index.html or 1-866-217-9197.
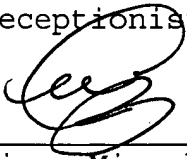
### *Conclusion*

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Diane D. Mizrahi whose telephone number is 571-272-4079.  The examiner can normally be reached on Monday-Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeffrey Gaffin can be reached on (571) 272-4146.  The fax phone numbers for the organization where this application or proceeding is assigned

are (703) 872-9306 for regular communications and (703) 305-3900

for After Final communications.

Any inquiry of a general nature or relating to the status

of this application or proceeding should be directed to the

receptionist whose telephone number is (703) 305-3900.

Diane Mizrahi
Primary Patent Examiner
Technology Center 2100

August 31, 2006